**Presented by**

Zahraa Alaaeddine

**2022-2023**

*Sentiment analysis*

Contents

[Abstract: 2](#_Toc124702432)

[ملخص: 4](#_Toc124702433)

[**Introduction** 5](#_Toc124702434)

[**Definition:** 6](#_Toc124702435)

[**Benefits of sentiment analysis** (kanade, 2022) 6](#_Toc124702436)

[**Tools of sentiment analysis** (kanade, 2022) 7](#_Toc124702437)

[**Awario (Bredava, 2022)** 8](#_Toc124702438)

[**Challenges of sentiment analysis** (repustate data in sight, 2022) 8](#_Toc124702439)

[**Challenges in informal Arabic language** (Al-Osaimi, Salha & Muhammad Badruddin Khan, 2017) 11](#_Toc124702440)

[**How does sentiment analysis works?** (spencer, 2022) 14](#_Toc124702441)

[Conclusion 16](#_Toc124702442)

[**References** 17](#_Toc124702443)

**Table of tables**

[**Table 1 DIFFERENT WORDS WITH THE SAME ROOT** 14](#_Toc124703174)

[**Table 2 repeated letters in the text** 14](#_Toc124703175)

[Table 3 negation words 14](#_Toc124703176)

# **Abstract**

Recently, there are wide numbers of users that use the social network like Twitter, Facebook, Instagram to share various kinds of resources, express their opinions, thoughts, messages in real time. Thus, increase the amount of electronic content that generated by users. Sentiment analysis becomes a very interesting topic in research community. It is a natural language processing (NLP) technique used to determine whether data is positive, negative or neutral. It is one of the most active research areas in natural language processing and text mining in recent years. It has been applied in a variety of domains including business, market research, manufacturing, etc. This research was set down to discuss sentiment analysis including its benefits, tools, challenges and how does it work.

**Keywords:** sentiment analysis, opinion mining, natural language processing (NLP), machine learning.

# **ملخص**

في الآونة الأخيرة، هناك عدد كبير من المستخدمين الذين يستخدمون الشبكة الاجتماعية مثل فايسبوك، تويتر، انستغرام للمشاركة أنواع مختلفة من الموارد والتعبير عن آرائهم وأفكارهم، الرسائل في الوقت الحقيقي. وبالتالي زيادة كمية الالكترونية المحتوى الذي تم إنشاؤه بواسطة المستخدمين. يصبح تحليل المشاعر موضوع مثير للاهتمام للغاية في مجتمع البحث. معالجة لغة طبيعية. هو الدراسة الحسابية لآراء الناس ومواقفهم ومشاعرهم المعبّر عنها بلغة مكتوبة. إنها واحدة من أكثر مجالات البحث نشاطًا في معالجة اللغة الطبيعية وتعدين النصوص في السنوات الأخيرة. تم تطبيقه في مجموعة متنوعة من المجالات بما في ذلك الأعمال التجارية، وأبحاث السوق، والتصنيع، وما إلى ذلك. وقد تم إعداد هذا البحث لمناقشة تحليل المشاعر بما في ذلك فوائده وأدواته وتحدياته وكيف يعمل.

**الكلمات المفتاحية:** تحليل المشاعر، تعدين الآراء، معالجة اللغات الطبيعية، التعلم الاّلي.

# **Introduction**

The opinions of others have a significant influence on daily decision-making-process. Accordingly, decisions taken will affect various aspects of our daily life. Before the Internet, people would seek opinions on products and services from sources such as friends, relatives, or consumer reports. However, For the past 30 years, the world has become increasingly more connected like never before. This means that opinions have become easier to share and are viewed by many more people, faster. So, people look to review sites (e.g., CNET, Epinions.com), e-commerce sites (e.g., Amazon, eBay), online opinion sites (e.g., TripAdvisor, Rotten Tomatoes, Yelp) and social media (e.g., Facebook, Twitter) to get feedback on how a particular product or service may be perceived in the market (Katrekar, n.d., p. 3). Importantly, these various opinions can be an extremely valuable resource for companies looking for insight into their performance, bottom line, this is where sentiment analysis come into play.

Sentiment analysis is a process of extensive exploration of data stored on the web to categorize the views expressed in a part of text written by consumers, usually in the form of reviews, a social media post, or employee correspondence. As a result, Sentiment analysis has been an active research field of natural language processing (NLP). It has also been widely studied in data mining, web mining, and information retrieval. It is an important factor when it comes to product and brand recognition, customer loyalty, customer satisfaction, advertising and promotion's success, and product acceptance. Thereupon, understanding the psychology of consumers can help product managers and customer success managers to alter their product roadmap with greater precision, so then increasing product and brand recall.

Sentiment analysis is originated from computer science. However, it has spread to management sciences and social sciences because of its importance to business and society as a whole. It is thus no surprise that sentiment analysis offers numerous challenging and fascinating solution to research problem that had never been attempted, and it’s not hard to imagine that sentiment analysis using social media might profoundly change the direction of research and practice in various field.

This study serves as an up-to-date and introductory text as well as comprehensive survey of sentiment analysis. Internally, it introduces sentiment analysis and its benefits and tools. Getting to the challenges that sentiment analysis is facing and solutions was suggested.

# **Definition:**

**Sentiment analysis** also referred to as opinion mining **is a field of study that uses computational methods to analyze, process, and reveal people’s feelings, sentiments, and emotions hidden behind a text or interaction. It uses machine learning (ML), natural language processing (NLP), data mining, and artificial intelligence (AI)techniques to mine, extract and categorize users’ opinions on a company, product, person, service, event, or idea for various sentiments.**

# **Benefits of sentiment analysis** (kanade, 2022)

**Offers integrated analysis**

sentiment analysis is linked to user behavior and it is further associated with demographics. It also highlights a specific emotional aspect. For example, corporate companies can use employee data of individuals who have left the organization to understand their feelings toward their colleagues, managers, and the company. This allows them to understand and correlate the similarities in the employee profiles that have raised the attrition issue. Moreover, the company can use integrated sentiment analysis to make changes in the company’s culture, employee policies, and constraints to lower the attrition rate.

**Efficient brand monitoring practice**

social media data is used to determine customer response to a product or service. With this, organizations can focus on improving a product that is most liked and preferred by all and to study how their products are performing in light of their competitors’ products.

**Product or service marketing**

With the launch of a new product, companies can employ sentiment analysis to understand user response to the new product rather than putting efforts into conducting surveys for the same. Customer reviews and ratings observed on ecommerce sites such as Flipkart and Amazon are an efficient way of flagging positive and negative sentiments toward a product.

Based on customer feedback, companies can zero in on speeding up the product production process, identify the features that need to be added.

**Efficient data mining practice**

Businesses can use sentiment analysis as a data mining tool that can help them gather competitive intelligence concerning competitor brands, thereby contributing to the business’s competitive advantage. [Neural networks](https://www.spiceworks.com/tech/artificial-intelligence/articles/what-is-a-neural-network/) play a crucial role in such data mining processes.

For example, brands can monitor and keep track of their social media mentions (social listening) with comments about their rivals. Then the analysis of such data gives an idea of consumer sentiments toward the products of competing brands. With such comparative data, companies can gain a competitive edge over other brands, allowing them to tweak or adjust their business model based on market sentiments.

**Supports political analysis**

Sentiment analysis on social media platforms such as Twitter can allow official authorities to keep a check on people’s reactions to newly-framed political policies. Political parties can reframe their policies and plan their election campaigns based on people’s responses, anger, and common trends.

Similarly, opposition parties can monitor public support for new laws and then use them to define their agendas. Moreover, the party or leader that is highly probable to sweep to victory in the elections can be forecasted through sentiment analysis.

# **Tools of sentiment analysis** (kanade, 2022)

Sentiment analysis tools refer to applications that dive deeper into user interactions to extract the intent, emotion, and tone behind them and derive more context from user conversations. Such tools come in handy for companies that use social media platforms, chats, or emails to engage with their consumers.

**Talk walker**

Talk walker's "Quick Search" is a sentiment analysis tool that's part of a larger customer service platform. It analyzes user comments, interactions, brand mentions, blogs, and chats to record their reactions to the activities and posts on social media. Thus, It uses[artificial intelligence](https://www.spiceworks.com/tech/artificial-intelligence/articles/what-is-artificial-intelligence/) to study user sentiments.

**Lexalytics**

Lexalytics is a unique sentiment analysis tool capable of giving insights into why a user responds to the provided service in a certain way. It uses NLP in the initial stages to process text and then performs the task of sentiment analysis to identify the emotions hidden in a user’s interaction. Therefore, Lexalytics concludes the process by compiling the information it derives into an easy-to-read and shareable display. Fundamentally, while most sentiment analysis tools offer insights into how users feel about something, Lexalytics tool answers the question ‘why’.

**Brandwatch**

Brandwatch is a popular sentiment analysis tool that keeps track of various social media aspects to reveal the user sentiment towards a service or brand.

One of the coolest features that Brandwatch provides is its "image insights" tool. which allows you to consider an image as input data in addition to text data. Moreover, it permits you to monitor logos shared over various digital platforms. This tool allows you to watch platforms where your brand images surface. You can also maintain a record of your brand’s performance for a specific target audience based on the customers’ emotions, tones, and attitudes.

**Awario (Bredava, 2022)**

Awario is a web-based social listening tool, with sentiment analysis being only a part of its vast capabilities. The data Awario analysis comes from social media platforms (including tweets, posts, etc.), forums, blogs, and websites. In addition to that, you can get access to sentiment analysis as soon as you log into the tool. For instance, gathering online data by putting the keywords you want to monitor. This can be related brand, product, competitors, industry, or any other phenomena being researched online.

# **Challenges of sentiment analysis** (repustate data in sight, 2022)

**Tone**

**problem**

Tone can be difficult to be interpreted verbally, and even more difficult to figure out in the written word. It even more complicated when one tries to analyze a massive volume of data that can contain both subjective and objective responses. That’s why brands can face difficulties in finding subjective sentiments and properly analyzing them for their intended tone.

**Solution**

The basis of any good sentiment analysis software includes the ability to decipher subjective statements from objective ones and then find the right tone in it. For example: “The product is gorgeous but not at that price” is a subjective sentiment but with a tonality that says that the price makes the product less attractive. With a smart sentiment API, companies can decipher such nuances in tone, at scale.

**Sarcasm**

**Problem**

People use irony and sarcasm in casual conversations and memes on social media. The act of expressing negative sentiment using backhanded compliments can make it difficult for sentiment analysis tools to detect the true context of what the response is actually implying. This can often result in a higher volume of “positive” feedback that is actually negative. A good sentiment analysis tool has to be able to detect sarcasm from the broader context. Otherwise, you’ll get inaccurate data about your brand at the end of the analysis.

**Solution**

A top-tier sentiment analysis API will be able to detect the context of the language used and everything else involved in creating actual sentiment when a person posts something. To this end , a good sentiment analysis tool has to be able to detect sarcasm from the broader context. Otherwise, you’ll get inaccurate data about your brand at the end of the analysis.

**Negations**

**Problem**

Negations, given by words such as not, never, cannot, were not, etc. can confuse the ML model. For example, a machine algorithm needs to understand that a phrase that says, “I can’t not go to my class reunion”, means that the person intends to go to the class reunion.

**Solution**

A sentiment analysis platform has to be trained to understand that double negatives outweigh each other and turn a sentence into a positive. This can only be done when there is enough corpus to train the algorithm and it has the maximum number of negation words possible to make the optimum number of permutations and combinations.

**Comparative sentences**

**Problem**

Comparative sentences can be tricky because they may not always give an opinion. Much of it has to be deduced. For example, when somebody writes, “the Galaxy S20 is larger than the Apple iphone12”, the sentence does not mention any negative or positive emotion but rather it states a relative ordering in terms of the size of the two phones.

**Solution**

Sentiment analysis accuracy can be achieved in this case when a sentiment model can compare the extent to which an entity has one property to a greater or lesser extent than another property. And then it ties that to a negative or positive sentiment. It is important to mention that it is not an issue of simply having a corpus of negative or positive sentiment-specific words, but in training the artificial intelligence machine to actually pull together information from its knowledge graph and analyze the relationship between entities, words, and emotions.

**Polarity**

**Problem**

Words such as “love” and “hate” are high on positive (+1) and negative (-1) scores in polarity. These are easy to understand. But there are in-between conjugations of words such as “not so bad” that can mean “average” and hence lie in mid-polarity (-75). Sometimes phrases like these get left out, which dilutes the sentiment score.

**Solution**

Sentiment analysis tools can easily figure out these mid-polar phrases and words in order to give a holistic view of a comment. In this context, a [topic-based sentiment analysis](https://www.repustate.com/topic-sentiment-analysis/) can give a well-rounded analysis, but with [aspect-based sentiment analysis](https://www.repustate.com/aspect-based-sentiment-analysis/), one can get an in-depth view of many aspects within a comment.

**Idioms**

**Problem**

Machine learning programs don’t necessarily understand a figure of speech. For example, an idiom like “not my cup of tea” will boggle the algorithm because it understands things in the literal sense. Hence, when an idiom is used in a comment or a review, the sentence can be misconstrued by the algorithm or even ignored. To overcome this problem a sentiment analysis platform needs to be trained in understanding idioms to manifold the problem.

**Solution**

The only way this challenge can be met with sentiment analysis accuracy is if the neural networks in an emotion mining API are trained to understand and interpret idioms. So, Idioms must be mapped according to nouns that denote emotions like anger, joy, determination, success, etc. Then these models are trained accordingly to give accurate insights from such text.

## **Challenges in informal Arabic language** (Al-Osaimi, Salha & Muhammad Badruddin Khan, 2017)

Arabic is a Semitic language spoken by more than 330 million people as a native language. Arabic is a highly structured and derivational language, in which morphology has a very important role. The Arabic natural learning processing(NLP)is fraught with many challenges, some of which result from the structural and morphological complexity of the language. Arabic is a derivational language, which means that many words can be formed from three-letter roots. The resulting words may look similar, but have very different meanings. Indeed, future research should consider the necessity of creating morphological analysis tools for Arabic text analysis that can cover all word forms and can perform suffix, affix, prefix, and root extraction. Grammatical analyzers and/or part-of-speech (POS) taggers are also needed. Some morphological analyzers have been developed for use with the Arabic language, such as BAMA (the Buckwalter Arabic Morphological Analyzer) and MADA (the Morphological Analysis and Disambiguation for Arabic analyzer).

There are no sophisticated POS taggers and lexicons tools in Arabic which identify all parts of speech and discover the difference of sentence's types. These issues present a challenge for sentiment mining, which generally requires both semantic analysis of words and grammatical analysis of text.

**major challenge that has surfaced due to the**

* **utilizing user’s own regional dialects**: emergence of social media is that most of the Arabic language found on the internet is written in informal Arabic. The informal version of the language is unstructured in nature. Furthermore, many users utilize their own regional dialects, rather than opting for modern standard Arabic; for instance, the word شوف /shoof, which means ―look‖ in English, might be used instead of the word أنظر/onthor.
* **Diacritics:** informal Arabic does not use diacritics; thus, in some cases, the meaning of the word becomes ambiguous. For example, the words مدرّسة―) teacher‖) and مدرسة(―school‖) look the same when written without diacritics (مدرسة). Also, when diacritics are removed and then add single space, so the word broken to many tokens.
* **same root for both informal and formal word:** Informal Arabic words usually do not have their own specific roots. Indeed, a stemmer will sometimes identify the same root for both the informal word and the formal word. table 1 exemplifies words with different meanings that share a common root.

**Table 1 DIFFERENT WORDS WITH THE SAME ROOT**

|  |  |  |  |
| --- | --- | --- | --- |
| Sentences | English translate | root | meaning |
| فش راحة بالبيت  Fesh ra7a bel bet | There is no comfort at home | روح | comfort |
| تروح عالملعب معي  troo7 3al mal3ab ma3e | Can you go to the playground with me | روح | go |

* **Repetition Letters**: It is known that in the Arabic language if we have repeated letters in the text it cannot occur more than twice. So if the repetition exists at beginning, middle or at the end of the word more than two times, it will be detected in the pre-processing step. Unfortunately, repetition cannot be detected where a letter is repeated only twice. table 2 exemplifies words with different meanings that share a common root.

**Table 2 repeated letters in the text**

|  |  |  |
| --- | --- | --- |
| sentences | English translate | After pre-processing |
| أحسسسسسسسسسسسن | **Better** | **أحسن** |
| بصراحة مرة واووووووووو | **In fact it is wow** | **بصراحة مرة واوو** |

* **Negations** The word polarities are affected significantly by ignoring negations like ما/Ma,لا /Laa, لم/lam, and لن/lan which are formal Arabic negations. The informal Arabic contains many of informal negation words like مو/Muo, مش/Mush, which also affect the text polarities by converting the meaning of the sentence to exactly the opposite. The informal Arabic used Arabized words. The Arabized words ―نو and ―نوت which means in English ―no‖ and ―not‖, are also used as negations words in informal Arabic

Table negation words

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| sentences | English translate | polarities | Sentence without negation | polarities |
| مو أهبل | **Not idiot** | **positive** | **أهبل** | **negative** |
| أنا مش طفل | **I am not a child** | **positive** | **أنا طفل** | **negative** |

* **Connecting different words together**: This problem involves usually Twitter users in connecting different words together; this method of writing occurs frequently in tweets because the length of a tweet is limited. they can cause ambiguities in meaning like words وفي/wa fee and وهم/wahem.
* In fact, the research on Arabic semantic is limited. One of those limited studies was provided by A. Shoukry and A. Rafea. They produce an application on Arabic sentiment analysis by classification the Arabic tweets. They used different ML classifiers and different features. They apply the SVM and naïve Bayes and also try the combinations of classifiers. Also, other researchers tried to find and explore the problems of sentiment analysis for informal Arabic. They apply their experiments on twitter. They use knowledge-based technique. There is a limitation in the number of Arabic sentiment lexicons, and the main challenge is to build lexicons for informal words

# **How does sentiment analysis works?** (spencer, 2022)

Sentiment analysis predominantly uses NLP and ML to make sense of the linguistic nuances observed in user interactions. The foundations of sentiment analysis are laid by the developers who design a[machine learning algorithm](https://www.spiceworks.com/tech/artificial-intelligence/articles/top-ml-algorithms/) capable of detecting content having varied sentiments.

**In the first step**, input data or text is gathered from social media, namely Twitter, using scraping tools, APIs, customers’ data feed, and from other reviews on services like Google and Yelp over a specific period of time.

**Second** text is cleaned where tools are allowed to process the data and prepare it for analysis by:

* Removing stop words (a, and, or, but, how, what…).
* Taking out punctuation (commas, periods…).
* Reducing words to their stem.

These tools allow to “clean” or “strip” the texts from irrelevant information to the analysis.

To this point, we can use our sentiment analysis algorithms to analyze the data we have gathered. Where input data or text is processed with the help of ML classifiers, which train large volumes of datasets that contain positive, negative, and neutral sentiments. In this phase, the data is divided into fundamental text components such as words, phrases, and sentences.

**In the training phase** at a typical ML classifier, input text goes through the feature extractor, which extracts features to generate feature vectors, labels, and tags (positive, negative, or neutral). Feature extraction methods based on word vectors give words with similar meanings a similar representation. The generated vectors are then inputted to the ML algorithm that produces a classifier model.

**Finally, in the prediction phase**, new data is fed into the feature extractor to generate feature vectors. Then classifier model processes to predict the sentiments hidden in the new input data. Classification algorithms such as Naïve Bayes,[linear regression](https://www.spiceworks.com/tech/artificial-intelligence/articles/what-is-linear-regression/), support vector machines, and deep learning are used to generate the output.

# **Conclusion**

Companies today make use of technologies that can help them not only to reach out to their customers wherever they are, but to also study and predict customers’ behavior and responses to products or services offered to them. sentiment analysis was represented as well as its (benefits, tools, challenges). However, this type of analysis can go through different fields and as a suggestion, where the government can use sentiment analysis to become more agile, efficient, and approachable and can take decisions to improve public services. As a result, citizens will not experience any downtime of services such as traffic jam or long wait at government service and thus maintaining a more positive view of their government.

# **References**

Al-Osaimi, Salha & Muhammad Badruddin Khan. (2017, february). Sentiment Analysis Challenges of Informal Arabic. *International Journal of Advanced Computer Science and Applications*, 280-283. Retrieved from file:///C:/Users/User/Downloads/Sentiment\_Analysis\_Challenges\_of\_Informal\_Arabic\_L.pdf

Bredava, A. (2022, may 19). *awario*. Retrieved from https://awario.com/blog/sentiment-analysis-tools/

kanade, v. (2022, august 12). *spiceworks*. Retrieved from https://www.spiceworks.com/tech/artificial-intelligence/articles/what-is-sentiment-analysis/

Katrekar, A. (n.d.). *an introduction to sentiment analysis*. Retrieved from globallogic.com: https://www.globallogic.com/se/wp-content/uploads/2019/12/Introduction-to-Sentiment-Analysis.pdf

Mayur Wankhade, Annavarapu Chandra Sekhara Rao & Chaitanya Kulkarni. (2022, february 7). A survey on sentiment analysis methods, applications, and challenges. *springer link*, p. 2.

*repustate data in sight*. (2022, june 11). Retrieved from https://www.repustate.com/blog/sentiment-analysis-challenges-with-solutions/

spencer, E. l. (2022, september 9). *revuze*. Retrieved from https://www.revuze.it/blog/sentiment-analysis/