e-ISBN : 978-629-98755-0-5

SIG : e-Learning@CS
https://appspenang.uitm.edu.my/sigcs/
Publication Date : 16-Oct-2023

SENTIMENT ANALYSIS USING MACHINE LEARNING ALGORITHM

*Syarifah Adilah Mohamed Yusoff[1], Nurul Izzah Mohd Rahiman[2], Wan Anisha Wan Mohammad[3]
Azlina Mohd Mydin[4] and Rozita Kadar[5]
**syarifah.adilah@uitm.edu.my[1], 2022755587@student.uitm.edu.my[2], wanan122@uitm.edu.my[3], azlin143@uitm.edu.my[4], rozita231@uitm.edu.my[5]*

[1,3,4,5]Jabatan Sains Komputer & Matematik (JSKM),
Universiti Teknologi MARA Cawangan Pulau Pinang, Malaysia

[2]Kolej Pengajian Komputer, Informatik dan Media,
Universiti Teknologi MARA Cawangan Kuala Terengganu, Malaysia

**Corresponding author*

ABSTRACT

In the current era of digitalization, technology has been deeply ingrained in society, serving as crucial tools not only for education but also significantly impacting the lifestyle of the community. One feature that emerges from social communications is the generation of diverse opinions and feedback across several domains, including books, people, hotels, products, research, events, and more. The phenomenon described, wherein feedback revolutionises the field, is commonly referred to as sentiment analysis, which has emerged as a captivating and burgeoning area of scholarly inquiry. This paper provides a foundational perspective and conceptual understanding of sentiment analysis, as well as the forthcoming challenges that this field will face.

Keywords: sentiment analysis, opinion mining, machine learning, challenges

Introduction

In the pre-internet era, individuals relied on many sources such as personal acquaintances, family members, consumer surveys, advertisements, and local communities to gather opinions regarding certain products and services. The proliferation of internet-based applications, driven by the widespread adoption of smartphones across all age groups, has significantly contributed to the expansion of social media platforms and blogs. Consequently, there has been an increase in the number of comments and evaluations pertaining to various daily activities.

Within a community, the opinions of others exert a substantial influence on our daily decision-making processes. These decisions encompass a wide range of choices, including the purchase of consumer goods like smartphones and the allocation of funds towards investments. Therefore, the perspectives can benefit organisations, governments, and individuals. One example involves the acquisition of consumer feedback pertaining to a marketing campaign, which enables an organisation to gauge the efficacy of said campaign or make necessary modifications to enhance its effectiveness. Meanwhile, collecting product feedback is advantageous for enhancing product quality, influencing income generation, and facilitating the evaluation of competitor goods (Hussein, 2018).

Sentiment analysis also known as Opinion Mining, a subfield within the discipline of natural language processing, focuses on the automated identification and classification of emotions and attitudes as expressed in written textual content. The procedure involves a number of crucial stages, which include preprocessing, feature extraction, and classification (Tan et al., 2023). Natural Language Processing (NLP) is a methodology that organises the extraction of various elements such as approaches, opinions, assessments, visions, perspectives, sentiments, emotions, feelings, excitements, and attitudes from a range of sources including verbal, nonverbal, Twitter, and databases. The process involves classifying opinions expressed in text as either "positive," "negative," or "neutral" (Balahur et al., 2010).

Level of Sentiment Analysis

Sentiment analysis can be categorised into multiple levels of analysis. Figure 1 shows the three main level of sentiment analysis.

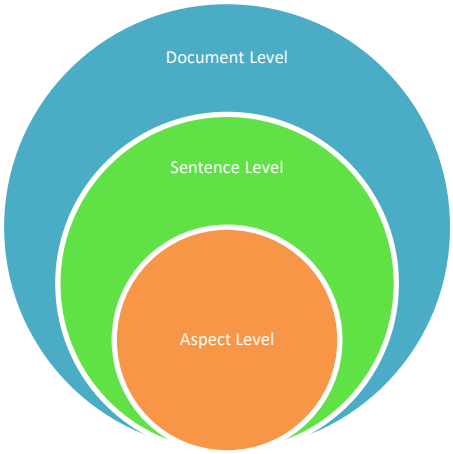

Figure 1: Three level of sentiment analysis

The task of document-level sentiment classification is a crucial issue within the field of sentiment analysis and opinion mining. It aims to evaluate the sentiment polarity of documents and online reviews (Shi et al., 2019). According to Rhanoui et al. (2019), when the quantity of words and noise words rises, the level of difficulty also increases, leading to a distortion in learning and complicating the prediction of polarity. The complete body of subjective text is considered the fundamental unit of information. The document is assumed to possess a viewpoint concerning a specific entity, such as a film, book, or hotel. This approach is not suitable when the text includes subjective viewpoints pertaining to many entities, such as non-forums and blogs. The entirety of the document is

categorised as either positive or bad. In order to facilitate the processing task, it is necessary to eliminate sentences that are not significant (Kolkur et al., 2015).

On the other hand, at the sentence level, the aim is to examine the statements and evaluate if they represented a positive, negative, or neutral attitude (Liu, 2012). The initial step is to determine if the sentence is subjective or objective in nature (Medhat et al., 2014a). Based on (Kolkur et al., 2015), the next step is the Sentiment Classification where this level will classify whether the sentence has positive, negative, or neutral sentiment. If the sentence is objective, then it is neutral, while if it the sentence is subjective, it is between positive or negative. The facts are contained in the objective sentence. It expresses no judgement or opinion about the item or thing, whereas subjective sentences express opinions.

Last but not least, aspect level refers to feature-based opinion mining and summarization. Classification is accomplished by finding and extracting product attributes from the source data (Patil & Yalagi, 2016). Aspect-level classification attempts to categorise sentiment in relation to certain properties of items (Medhat et al., 2014b). Based on the study from (Liu, 2012), aspect level examines the opinion itself rather than linguistic constructions. It is founded on the notion that an opinion is made up of a feeling (either positive or negative) and a target (of opinion). An opinion is only useful if its goal is recognised. There are two (2) types of opinions: regular and comparative opinions.

Approach used in Sentiment Analysis

There are three different approaches to deal with the sentiment analysis problem: (1) strategies based on lexicons, (2) techniques based on machine learning, and (3) hybrid approaches (Dang et al., 2020). Figure 2 shows that there are three types of machine learning approaches: unsupervised, semi-supervised, and supervised learning. Meanwhile, there are two types of Lexicon-based approaches: corpus and dictionary based.

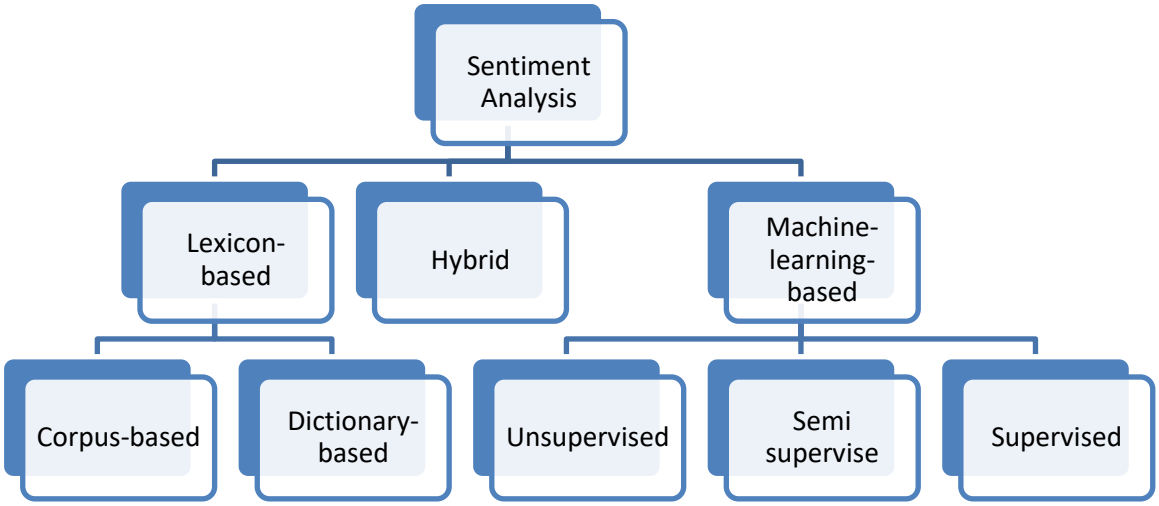

Figure 2: Sentiment Analysis Approaches

Implementation of Machine Learning for Sentiment Analysis

Work	**Problem**	**Objective**	**Algorithm Used**
Sentiment Analysis On E-Sports for Education Curriculum Using Naive Bayes and Support Vector Machine. (Ardianto et al., 2020)	The need to accurately analyse a large volume of comments on social media regarding e-sports education to separate positive and negative sentiments.	To measure opinions and separate positive and negative sentiments towards e-sports education	Naïve Bayes, Support Vector Machine (SVM)
Opinion Analysis for Emotional Classification on Emoji Tweets using the Naïve Bayes Algorithm. (Sendari et al., 2020)	Miscommunication can occur when Emojis are used separately within a message, leading to ambiguity of emotions conveyed.	The study aims to classify tweet content based on Emojis, identify emotions (e.g., anger, joy, sadness), and determine the dominant emotional trend on Twitter.	Naïve Bayes
Sentiment analysis of COVID-19 vaccine in Indonesia using Naïve Bayes Algorithm. (Pristiyono et al., 2021)	Concerns and uncertainties exist regarding public acceptance of COVID-19 vaccines in Indonesia.	To assess public opinion on the COVID-19 vaccine in Indonesia using social network	Naïve Bayes

Using Naïve Bayes Algorithm in detection of Hate Tweets. (Kiilu et al., 2018)	The presence of hateful speech on social media platforms, which can have negative impacts on users and incite violence.	To develop a reliable tool for detecting hate tweets and classifying hateful speech on social media.	Naïve Bayes
Sentiment Analysis of Government Policy on Corona Case Using Naive Bayes Algorithm. (Isnain et al., 2021)	Determining whether the sentiment expressed on Twitter regarding the policy is positive or negative.	To analyse public sentiment on Twitter about the New Normal policy and evaluate the performance of TF-IDF and N-gram using Naïve Bayes Classifier.	Naïve Bayes
App Review Sentiment Analysis Shopee Application in Google Play Store Using Naive Bayes Algorithm. (Pratmanto et al., 2020)	Analysing user sentiments towards the Shopee app based on Google Play Store comments.	Assisting Shopee's management in identifying positive or negative opinions and contributing to the development of related theories.	Naïve Bayes
Implementation of Sentiment Analysis on Twitter Using Naïve Bayes Algorithm to Know the People Responses to Debate of DKI Jakarta Governor Election. (Pratama et al., 2019)	Analysing vast social media data, like Twitter, poses challenges in extracting meaningful insights accurately.	To analyse Twitter sentiment on Jakarta's governor election debates, focusing on public response to three candidates.	Naïve Bayes
Sentiment Analysis of Tweets on Social Issues using Machine Learning Approach. (Kaur & Sharma, 2020)	Analysing the sentiments of people related to five major social issues (corruption, women violence, poverty, child abuse, illiteracy) using tweet data collected from 2006 to July 2020.	To analyse sentiments in tweets about social issues using machine learning and NLP techniques.	Decision Tree, Naïve Bayes, Random Forest (RF), Logistic Regression (LR), Support Vector Machine (SVM), K-nearest Neighbours (KNN), Shastic Gradient Descent (SGD)
Sentiment analysis for customer review: Case study of Traveloka. (Diekson et al., 2023)	Analysis of sentiments from Twitter users regarding the services provided by Traveloka.	To determine the sentiment of Twitter users regarding Traveloka's services and assess customer satisfaction.	Naïve Bayes, Logistic Regression (LR), Support Vector Machine (SVM)
Sentiment Analysis of YouTube Movie Trailer Comments Using Naïve	Analysis of viewers' comments and opinions about the	To conduct sentiment analysis on viewers'	Naïve Bayes

Bayes. (Novendri et al., 2020)	Money Heist series in order to classify their sentiments.	comments using the Naïve Bayes algorithm to determine the opinions and sentiments expressed about the Money Heist series.	

Challenge in Sentiment Analysis

Sentiment analysis in every aspect of social issues have discovered a lot of insight feedback came across from huge information circulate from online conversation. Anyhow, the complexity of human language constitute challenge for AI to interpret the real meaning of the sentence. For example, Figure 3 depicted a consumer feedback on a product of a handphone.

(1) I bought a Zfold3 Samsung handphone 3 months ago.
(2) I simply love it
(3) The camera quality is fantastic
(4) The flip screen is amazing for reading
(5) However, my husband thinks it too heavy to put in his pants pocket

Figure 3: Consumer feedback on a product

This feedback shows both positive and negative sentiment, where the first statement until forth indicate positive based on the word “love”, “fantastic” and “amazing”. On the fifth statement, the sentence indicates the negative word such as “too heavy”. Hence, how would AI decide the feedback is positive/negative/neutral.

The meaning of a word depends on the context it is used in. Sarcasm, polarity and polysemy are example of challenges happen in understanding the real intention of a sentence. Sarcasm refer to a way in expressing a thought in opposite way, example is when the intention is negative but the sentence is positive, example “I am soooooo happy the item shipped after 2 weeks”. In such case, the sentiment analysis tool can classify the sentence as positive, but in reality, it is negative. Polarity is used to identified the level of satisfaction, some feedback can be very apparent such as “It was terrified experience”. Other feedback “the experience is unexpected”, the level of satisfaction is unclear, infact the level of positive or negative satisfaction is unknown. Polysemy is a sentence that have multiple related meaning, “your customer support is killing me VS your product is killing me”. Then it become more challenging for algorithm to differentiate the intended meaning. Negation detection such as “no,

not, -non, -less, dis" does not mean the whole statement is negative, example is "it is not unhealthy". Some of the modern negation detection methods are not sufficient in detecting and classify the statement correctly and bound to make many errors (Simmering & Perry, 2023).

Multilingual data which is happen to be the most critical and complicated problems in every cultural post since English is most of spoken language and mix in every community spoken slang. Conflict with this limitation, some of the analysis might get lost in translation due to the sentiment tool are primary trained to categorized word in one language (Yilmaz, 2023). Example is Malay slang "Yesterday's concert was Best Gila!" the direct translation might be negative, but "Best Gila" is a slang used to say "awesome" and suppose to be a positive statement.

Sentiment analysis primarily relies on textual statements for determining emotions. However, emojis, which are considered out-of-vocabulary words, have become integrated into sentences as a means of expressing emotions and enhancing the accuracy of sentiment prediction. It is important to note that emojis may be subject to varying interpretations across different groups. Therefore, it is imperative for sentiment analysis technologies to possess the capability of incorporating emojis in order to enhance the accuracy of the analysis.

Conclusion

Sentiment analysis is regarded as a significant justification for comprehending the genuine needs and contentment of consumers, as well as a potent instrument for predicting the future viability of products and services. The utilisation of machine learning and Artificial Intelligence in addressing upcoming challenges serves to highlight the fact that sentiment analysis remains a comparatively underexplored area of research.

References:

Ardianto, R., Rivanie, T., Alkhalifi, Y., Nugraha, F. S., & Gata, W. (2020). Sentiment analysis on E-sports for education curriculum using naive Bayes and support vector machine. *Jurnal Ilmu Komputer Dan Informasi*, *13*(2), 109–122.

Bhadane, C., Dalal, H., & Doshi, H. (2015). Sentiment analysis: Measuring opinions. *Procedia Computer Science*, *45*(C), 808–814. https://doi.org/10.1016/j.procs.2015.03.159

Dang, N. C., Moreno-García, M. N., & de la Prieta, F. (2020). Sentiment analysis based on deep learning: A comparative study. *Electronics (Switzerland)*, *9*(3). https://doi.org/10.3390/electronics9030483

Diekson, Z. A., Prakoso, M. R. B., Putra, M. S. Q., Syaputra, M. S. A. F., Achmad, S., & Sutoyo, R. (2023). Sentiment analysis for customer review: Case study of Traveloka. *Procedia Computer Science*, *216*, 682–690.

Hussein, D. M. E. D. M. (2018). A survey on sentiment analysis challenges. *Journal of King Saud University - Engineering Sciences*, *30*(4), 330–338. https://doi.org/10.1016/j.jksues.2016.04.002

Isnain, A. R., Marga, N. S., & Alita, D. (2021). Sentiment Analysis Of Government Policy on Corona Case Using Naive Bayes Algorithm. *IJCCS (Indonesian Journal of Computing and Cybernetics Systems)*, *15*(1), 55. https://doi.org/10.22146/ijccs.60718

Kaur, C., & Sharma, A. (2020). Sentiment analysis of tweets on social issues using machine learning approach. *International Journal of Advanced Trends in Computer Science and Engineering*, *9 (4), 6303–6311. Https://Doi. Org/10.30534/Ijatcse/2020*, *310942020*.

Kiilu, K. K., Okeyo, G., Rimiru, R., & Ogada, K. (2018). Using Naïve Bayes Algorithm in detection of Hate Tweets. *International Journal of Scientific and Research Publications (IJSRP)*, *8*(3). https://doi.org/10.29322/ijsrp.8.3.2018.p7517

Kolkur, S., Dantal, G., & Mahe, R. (2015). *Study of Different Levels for Sentiment Analysis.*

Liu, B. (2012). *Sentiment Analysis and Opinion Mining*. Morgan & Claypool Publishers.

Medhat, W., Hassan, A., & Korashy, H. (2014). Sentiment analysis algorithms and applications: A survey. *Ain Shams Engineering Journal*, *5*(4), 1093–1113. https://doi.org/10.1016/j.asej.2014.04.011

Novendri, R., Callista, A. S., Pratama, D. N., & Puspita, C. E. (2020). Sentiment Analysis of YouTube Movie Trailer Comments Using Naïve Bayes. *Bulletin of Computer Science and Electrical Engineering*, *1*(1), 26–32. https://doi.org/10.25008/bcsee.v1i1.5

Patil, P., & Yalagi, P. (2016). *Sentiment Analysis Levels and Techniques: A Survey*.

Pratama, Y., Roberto Tampubolon, A., Diantri Sianturi, L., Diana Manalu, R., & Frietz Pangaribuan, D. (2019). Implementation of Sentiment Analysis on Twitter Using Naïve Bayes Algorithm to Know the People Responses to Debate of DKI Jakarta Governor Election. *Journal of Physics: Conference Series*, *1175*(1). https://doi.org/10.1088/1742-6596/1175/1/012102

Pratmanto, D., Rousyati, R., Wati, F. F., Widodo, A. E., Suleman, S., & Wijianto, R.

(2020). App Review Sentiment Analysis Shopee Application in Google Play Store Using Naive Bayes Algorithm. *Journal of Physics: Conference Series*, *1641*(1). https://doi.org/10.1088/1742-6596/1641/1/012043

Pristiyono, Ritonga, M., Ihsan, M. A. al, Anjar, A., & Rambe, F. H. (2021). Sentiment analysis of COVID-19 vaccine in Indonesia using Naïve Bayes Algorithm. *IOP Conference Series: Materials Science and Engineering*, *1088*(1), 012045. https://doi.org/10.1088/1757-899x/1088/1/012045

Rhanoui, M., Mikram, M., Yousfi, S., & Barzali, S. (2019). A CNN-BiLSTM Model for Document-Level Sentiment Analysis. *Machine Learning and Knowledge Extraction*, *1*(3), 832–847. https://doi.org/10.3390/make1030048

Sendari, S., Zaeni, I. A. E., Lestari, D. C., & Hariyadi, H. P. (2020). Opinion analysis for emotional classification on emoji tweets using the Naïve bayes algorithm. *Knowledge Engineering and Data Science*, *3*(1), 50–59.

Shi, T., Wang, S., Rakesh, V., & Reddy, C. K. (2019). Document-level multi-aspect sentiment classification for online reviews of medical experts. *International Conference on Information and Knowledge Management, Proceedings*, 2723–2731. https://doi.org/10.1145/3357384.3357828

Simmering, P., & Perry, T. (2023, February). *10 Challenges of sentiment analysis and how to overcome them Part 2*. ESOMAR. https://researchworld.com/articles/10-challenges-of-sentiment-analysis-and-how-to-overcome-them-part-2

Steinberger, R., Kabadjov, M., Zavarella, V., van der Goot, E., Halkia, M., Pouliquen, B., & Belyaeva, J. (2010). Sentiment analysis in the news. *Proceedings of the 7th International Conference on Language Resources and Evaluation, LREC 2010*, 2216–2220

Tan, K. L., Lee, C. P., & Lim, K. M. (2023). A Survey of Sentiment Analysis: Approaches, Datasets, and Future Research. In *Applied Sciences (Switzerland)* (Vol. 13, Issue 7). MDPI. https://doi.org/10.3390/app13074550

Yilmaz, B. (2023, September 8). *Top 5 Sentiment Analysis Challenges and Solution in 2023*. AIMultiple. https://research.aimultiple.com/sentiment-analysis-challenges/