



## Sentiment Analysis of top colleges using social media data: Systematic Literature Review.

R. V. Patil<sup>#1</sup>, Mayur Tembhare<sup>#2</sup>, Asha Mandale<sup>#3</sup>, Bhushan Asati<sup>#4</sup>, Ajesh Rathod<sup>#5</sup>,

<sup>#</sup>Computer Engineering Department, SavitriPhule University  
Manjari (BK), Pune - 412307(Maharashtra) India

<sup>1</sup>[rvpatil3475@yahoo.com](mailto:rvpatil3475@yahoo.com)

<sup>1</sup>[tembharemayur@gmail.com](mailto:tembharemayur@gmail.com)

<sup>2</sup>[ashamandale28@gmail.com](mailto:ashamandale28@gmail.com)

<sup>3</sup>[bhushanasati25@gmail.com](mailto:bhushanasati25@gmail.com)

<sup>4</sup>[ajeshrathod2000@gmail.com](mailto:ajeshrathod2000@gmail.com)

**Abstract**— This paper is a report on the sentiment analysis in social media that explored the methods, social media platforms used and applications specifically finding the Top Colleges across the globe. Social media contain a large amount of raw data that has been uploaded by users in the form of text, videos, photos, and audio. The data can be converted into valuable information by using Sentiment analysis. A systematic review of the studies published between 2014 to 2021 was undertaken using the following trusted and credible databases including ACM, Emerald Insights, IEEE Xplore, Science Direct, and Scopus. After the initial and in-depth screening of the paper, 14 out of 86 Articles / Research papers have been chosen from the review process. The article has been reviewed based on the aim of the study. The result shows most of the articles applied the opinion-lexicon method to analyze text sentiment in social media, extracted data on microbiology sites mainly Twitter, and sentiment analysis applications can be seen in world events, healthcare, politics, and business.

**Keywords**— Sentiment analysis, Big Data, Social media, College Ranking.

### I. INTRODUCTION

The emergence of web 2.0 and the introduction of web 3.0 is changing the world of social media. Not only is online social media used to connect and share information and personal opinion with others, but even businesses can also communicate, understand and improve their products and services through connecting in social media. The number of social media users increases every day and it is estimated in 2022 is 4.55 billion social media users worldwide which is 9.9 % higher than last year worldwide [1]. There is various type of information uploaded and shared on social media in the form of text, videos, photos, and audio [2]. Social media is rich with raw and unprocessed data and the improvement in technology, especially in machine learning and artificial intelligence, allows the data to be processed and converted into useful data that they can benefit most business organizations [3].

This paper focuses to provide a better understanding of the application of sentiment analysis in social media platforms by examining related literature published between 2014 to 2019. Sentiment analysis is an approach that uses Natural Language Processing (NLP) to extract, convert and interpret opinions from a text and classify them into positive, negative, or natural sentiments [11]. Most of the previous

studies applied sentiment analysis to a product or movie review to better understand their customer and make the necessary decision to improve their product or services [12].

Scholars have been conducting a study on sentiment analysis since the last decade when most papers started to appear and rapidly grew after the year 2004 [13]. Sentiment analysis is divided into three different levels which are sentence level, document level, and feature level. The purpose is to classify the opinion either from sentence, document, or features into positive and negative sentiment [14].

There are 2 main methods of sentiment analysis have been identified which are a machine-learning approach and a lexicon-based approach. The machine learning approach utilized algorithms to extract and detect sentiment from data while the lexicon-based approach works by counting the positive and negative words that are related to the data. Scholars have been developing a new effective and accurate model in sentiment analysis. But there a challenge arises in developing a model where most of it is designed for the English language. But a recent study shows that there is a sentiment analysis model design in other languages such as Korean [15], Thailand [16], Arabic [17], Malay [18], Portuguese [19], and Chinese [20]. As for the application of sentiment analysis, it is reported that it has been done in business and marketing, politics, and public action context. Example of application is E-commerce, voting application, and world events [21]. Most of the data extracted for the study was extracted from social media. Social media contain a vast amount of data from online users and we can get any information on a product, service, place, or event which makes it fit for sentiment analysis study.

### II. THE REVIEW

#### A. Design:

A systematic review was undertaken using 6 steps guidelines for conducting a systematic literature review in management [22]. First, we start by defining the research question. Then determine the required characteristic for the study. Continuing by retrieving potentially relevant literature and selecting pertinent literature. We then synthesize relevant information from the literature and the final step is reporting the result of the review.

B. To provide an overview of the review, the following research question was addressed:

- RQ1: What is the method used in sentiment analysis of social media?
- RQ2: What is the type of social media platform used to apply sentiment analysis?
- RQ3: What is the application context of sentiment analysis in social media?
- RQ4: How it can be utilized to find the Best College in Top Major Cities?

C. Retrieving and selecting pertinent literature:

The review utilized five reputable and credible online databases that published literature covering information and computer science area. Search strings keywords used for all five online databases are "Sentiment analysis, social media, Facebook, Twitter". The total number of articles identified from the database search is 395 articles. 34 articles were identified from Emerald Insight, 216 results were identified by science direct, 30 results from Association for computing machinery (ACM), 64 articles from Scopus, and 51 articles were identified by IEEE. Then the screening of papers is conducted based on the inclusion and exclusion criteria and the screening resulted in 145 articles. Consequently, the screening involved reading the full texts and analyzing each article and we obtain 14 finalized articles.

D. Synthesizing the literature:

Author	Title	Method / Tools	Application / Result	Context
Fulyaman, Djama & Sukoco. (2017)	Sentiment Mining of Community Development Program Evaluation Based on Social Media	Lexicon-based and machine learning	Success level of the community development program	Twitter
Martín-Domingo, Marina, & Mmadiberg. (2019)	Social media as a resource for sentiment analysis of Airport Service Quality	Machine learning	Analyse airport service quality	Twitter account
Mansour. (2018)	Social Media Analysis of User's Responses to terrorism using sentiment analysis and text mining	Lexicon-based	Most user view ISIS as a threat and fear	Twitter
Saragih & Girsang. (2017)	Sentiment Analysis of Customer Engagement on Social Media in Transport Online	Lexicon-based	Evaluate the business performance of online transport.	Facebook and Twitter comments
Hassan, Hussain, Husain, Sadiq, Lee. (2017)	Sentiment Analysis of Social Networking Sites (SNS) Data using Machine Learning Approach for the Measurement of Depression	Machine learning	Find the depression level of a person	Twitter and newsgroup
Joyce & Deng. (2017)	Sentiment Analysis of Tweets for the 2016 US Presidential Election	Lexicon-based and machine learning	Calculate sentiment expressed and compare with polling data to see the correlation	Twitter
Boro, Harmina, Malik, & Batista-Navarro. (2018)	Analyzing Sentiments Expressed on Twitter by UK Energy Company Consumers	Lexicon-based	Analyze energy provider company and the sentiment that users show	Twitter
Hao & Dai. (2016)	Social media content and sentiment analysis on consumer security breaches	Lexicon-based	Security breaches can be detected in the early stages and prevent further destruction	Twitter
Shayan, Wai, Ching, Sulaiman, Jafar & Zakaria. (2017)	Social Media Sentiment Analysis on Employment in Malaysia	Lexicon-based	Negative sentiment score on employment	Multiple channel social media
Ish, Trudie & Neagu. (2014)	Social Media Analysis for Product Safety using Text Mining and Sentiment Analysis	Lexicon-based and machine learning	Monitor brands in order to act in even of a sudden rise in negative sentiment.	Facebook comment and Twitter
Ali, Dong, Bougattaya, Erradi & Hadjidi. (2017)	Sentiment Analysis as a Service: A social media-based sentiment analysis framework	Machine Learning	Identify the location of disease outbreaks	Twitter, Reddit, Instagram, news forum.
Akter, Aziz & Tareq. (2016)	Sentiment analysis on facebook group using the lexicon-based approach	Machine learning	Determine recent trends and characteristics of people food habit.	Facebook Group - Foodbank
Mahnab, Islam & Rahman. (2018)	Sentiment Analysis on Bangladesh Cricket with Support Vector Machine	Lexicon-based and machine learning	Analyze people sentiment expressed towards cricket	Facebook Group - Bangladesh Cricket
Cbedia Cynthia & Tau (2017)	Social media sentiment analysis: lexicon versus machine learning	Lexicon-based and Machine learning	Sentiment analysis on consumer generated content	Facebook brand pages

### III. REPORTING THE RESULT

A. The Sentiment analysis method used in social media:

Based on the papers reviewed, all of the papers demonstrated the usage of either the Lexicon-based method, the Machine learning method, or a mix of both methods when implementing sentiment analysis. The results show in conducting sentiment analysis, 7 of the reviewed paper uses the lexicon-based method, 10 papers use machine learning and 7 papers show the combination of both methods.

The lexicon-based method is known as an unsupervised learning method. The lexicon method does not require any training data and only depends on the dictionary. Most of the study adapted

Sentiwordnet and TF-IDR methods when conducting sentiment analysis. This approach is calculated based on the occurrences of the terms in the text data with other positive or negative words in the predeveloped polarity lexicons like Sentiwordnet [11]. The TF-IDR method works by converting the words into a number and it is calculated using the term frequency-inverse document frequency method [23].

The techniques rely on lexical resources and the effectiveness of the whole approach strongly depends on the quality of the lexical resources. It is based on the polarity of a piece of text that can be obtained on the ground of the polarity of the words which compose it. Due to the complexity of natural languages, this approach is not designed to cover all aspects of language especially when it comes to slang, sarcasm, and negation [24]. Using sentiment words is not sufficient. Some of the problems exist such as some words having different meanings based on the application, some sentences containing sentiment words may not express any sentiment and many sentences without sentiment words can also imply opinion [18]. However, the lexicon-based method does have its own advantage such as it provides simple counting of positive and negative words, flexible to fit with different languages, and speed to complete analysis. The machine learning method falls under supervise learning and the method requires training data in order to be processed. The most used method in machine learning method is the SVM and Naïve Bayes model. Different machine learning models but these are the most commonly used. Naïve Bayes is successful when applied to well-formed text corpus [25] while support vector machine gives a good performance for low-shape datasets. Nevertheless, the machine learning method performs poorly on Facebook with people posting in random lengths and lots of spelling mistakes and it requires a huge amount of training samples in order to adapt the method as the amount of dataset will influence the size and quality of the output [26, 27]. Furthermore, analyzing with machine learning is time-consuming and it takes hours in the complex machine learning model especially if training is required [18]. The process is faster with a smaller size training dataset but it leads to poorer classification accuracy [28].

Interestingly, researchers argue that both types of analysis methods perform very similarly in terms of accuracy [28]. There are options to combine two approaches mainly lexicon-based sentiment classification that contain sentiment scoring function and Naïve Bayes multinomial event models from a machine learning approach to predict the direction of sentiment. Instead of relying on one method, studies have proven that combining both methods has better efficiency [19]. Thus, in order to improve the outcome, it is recommended to combine both methods as they will complement each other, and the result is improved compared to using one approach only. A combined approach is valuable to identify a phenomenon [28]. It also can improve the handling of unstructured data [29].

B. Types of social media platforms use to extract data for Sentiment analysis:

With the usage of API, people can access and copy the data on any desired topic based on keywords or hashtags. Twitter conducts real-time analysis and closely public sentiment as Twitter has about 500 million tweets per day and around 200 billion tweets per year and it allows public access to its data through API [3]. Twitter is used to search and collect tweets from 8 different countries from western and eastern countries. There is Twitter user all around the world thus making it rich with opinion and views by people from a different country, different language, and different perception [4].

For example, Twitter is used to collect users' tweets on particular presidential candidates during elections [5] and collected tweets that had been written on a community development program activity [6].

Moreover, Twitter is also used to collect messages from customers to energy companies in the UK [7] and analyze tweets downloaded from London Heathrow airport's official Twitter account to be analyzed further using sentiment analysis [8].

Facebook has the largest number of social media users in the world. But it is not very popular for sentiment analysis as the data is messy, it is not structured well, and people often use short forms and a lot of spelling errors. This makes the data harder to be analyzed. An example

is using Facebook and Twitter to fetch pages, status updates, and comments suggesting user experiences [9].

A study conducted gathered data from various sources of social media including forums, blogs, Expedia, blog spots, mainstream media, WordPress, YouTube, Twitter, aggregator, and Facebook. And the result shows that 88% of the data comes from Twitter [10].

The other source of social media is not preferable because the number of data or opinions that can be extracted is limited such as in Blogspot, YouTube, and WordPress.

### C. Application Context of Sentiment Analysis:

The application of sentiment analysis ranges from business and marketing, politic, and health to public action. Sentiment analysis is not limited to one application, but it provides a vast application in different areas to assist in decision-making. Sentiment analysis can be applied to world events such as an event, activity, sports, or disaster that is occurring in world [27, 30]. Some of the examples are a study conducted to compare how people from western countries and eastern countries view ISIS. The result shows how two sides of the world view ISIS the same way which is a terrorist [4]. Sentiment analysis also allows for raising awareness of data security and the danger of security breaches. It also acts as a guideline for companies to respond to security breaches in shaping public perception [24]. Furthermore, sentiment analysis was also conducted on the unemployment rate and employment sentiment score in social media [31].

We can see the application of sentiment analysis in healthcare and where the study uses Sentiment analysis as a service framework proposed and utilize Spatio-temporal properties to identify locations of disease outbreaks [32]. In addition, sentiment analysis can identify the sentiment needs of people during a disaster and prepare an appropriate response to rescue [33]. Moreover, Sentiment analysis allows finding the level of depression of a person by overserving and analyzing emotions from text [25].

Sentiment analysis can be used to predict political election where it shows the data analyzed from Twitter is more reliable as a platform where 94% of correlation has been found to polling data and have the potential to become a platform that is able to rival sophisticated polling techniques [34].

### D. Utilization of Sentiment analysis For Finding Best Colleges in Particular Cities:

- I. Sorting Data at scale -sentiment analysis helps business process huge amounts of unstructured data in an efficient and cost effective way.
- II. Real-time Analysis -sentiment analysis can identify critical issues in real time
- III. Consistent criteria -Tagging text by sentiment is highly influenced .It is estimated that people only agree of the time of sentiment of particular text.

### IV. CONCLUSIONS

The conducted systematic literature review provides information on studies on sentiment analysis in social media. The paper makes the following three contributions. First, we show what is the method used in analyzing sentiment in social media. There is various method introduced by researchers, still, the most common method uses in Lexicon-based method is SentiWordnet and TF-IDF while for machine learning is Naïve Bayes and SVM. Choosing the appropriate method of sentiment analysis is depending on the data itself. Both methods demonstrated a similar accuracy. The things that need to take into consideration are the structure of the text, time, and amount of data. If the data structure is messy, with a small amount of data and limited time available for analyses, it is recommended to go for the lexicon-based method. Bigger data is suitable for machine learning-based methods as it requires more time and data to train. In order to improve the quality and accuracy of the result, it is suggested to combine both the lexicon and machine learning methods.

Second, we identify what is the most common type of social media site to extract information for sentiment analysis. The most popular social media site to extract information is Twitter. Most of the reviewed papers use Twitter as their social media context. This is due to the availability, accessibility, and richness of Twitter content. There are millions of tweets every day on almost any topic. This indicates that social media is becoming a precious source of information. However, less attention is given to other social media sources such as blogs, WordPress, YouTube, and others. The content of each social media might be different, and it is worth exploring other sources that might open to new knowledge and findings.

Third, we demonstrate the application of sentiment analysis in social media. Sentiment analysis has a broad application and can be utilized in different areas such as improving quality and strategy in business, political forecasting an election result, monitoring disease outbreaks, creating awareness on the importance of data security, perception towards a particular sport, and improving locate and respond to the disaster. This shows that sentiment analysis plays a huge role to understand people's perceptions and helps in decision-making. For future recommendation, further investigation is needed to develop a universal model of sentiment analysis that can be applied to a different type of data, explores other potential social networking sites to obtain users' opinion, and expand the context of sentiment analysis application.

### V. REFERENCES

- [1]. Kemp, S. (2022, February 1). Social media users pass the 4.5 billion Mark. We Are Social UK. Retrieved November 5, 2022, from <https://wearesocial.com/uk/blog/2021/10/social-media-users-pass-the-4-5-billion-mark/>
- [2]. Giri, Kaiser J, and Towseef A Lone. (2014). "Big Data-Overview and Challenges." International Journal of Advanced Research in Computer Science and Software Engineering 4 (6).
- [3]. Sayce, D. (2022, August 16). The number of tweets per day in 2022. David Sayce. Retrieved November 5, 2022, from <https://www.dsayce.com/social-media/tweets-day/#:~:text=Every%20second%2C%20on%20average%2C%20around%206%2C000%20tweets%20are%20tweeted%20on,200%20billion%20tweets%20per%20year.>
- [4]. Mansour, Samah. (2018) "Social Media Analysis of User's Responses to terrorism using sentiment analysis and text mining." Procedia Computer Science 140: 95–103.
- [5]. Makazhanov, Aibek, and Davood Rafiei. "Predicting political preference of Twitter users." Proceedings of the 2013 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining. 2013.
- [6]. Yuliyanti, Siti, Djatna, Sukoco Taufik, and Heru. (2017) "Sentiment Mining of Community Development Program Evaluation Based on Social Media." TELKOMNIKA (Telecommunication Computing Electronics and Control) 15 (4): 1858-1864.
- [7]. Ikoro, Victoria, Maria Sharmina, Khaleel Malik, and Riza Batista-Navarro. (2018) "Analyzing Sentiments Expressed on Twitter by UK Energy Company Consumers", in Fifth International Conference on Social Networks Analysis, Management and Security (SNAMS) (pp. 95-98): IEEE.
- [8]. Martin-Domingo, Luis, Juan Carlos Martin, and Glen Mandsberg. (2019) "Social Media as a Resource for Sentiment Analysis of Airport Service Quality (ASQ)." Journal of Air Transport Management.
- [9]. Isah, Haruna, Paul Trundle, and Daniel Neagu. (2014) "Social Media Analysis for Product Safety Using Text Mining and Sentiment Analysis", in 14th UK Workshop on Computational Intelligence (UKCI): IEEE.
- [10]. Shayaa, Shahid, Phoong Seuk Wai, Yeong Wai Chung, Ainin Sulaiman, Noor Ismawati Jaafar, and Shamshul Bahri Zakaria. (2017) "Social Media Sentiment Analysis on Employment in Malaysia", in

the Proceedings of 8th Global Business and Finance Research Conference, Taipei, Taiwan.

- [11]. Agarwal, Basant, Namita Mittal, Pooja Bansal, and Sonal Garg. (2015) "Sentiment Analysis Using Common-Sense and Context Information." *Journal of Computational Intelligence and Neuroscience* 9 (2015).
- [12]. U. T. Gursoy, D. Bulut, and C. Yigit. (2017) "Social Media Mining and Sentiment Analysis for Brand Management." *Global Journal of Emerging Trends in e-Business, Marketing and Consumer Psychology* 3 (1): 497-551.
- [13]. Mäntylä, Mika V., Daniel Graziotin, and Miikka Kuutila. (2018) "The Evolution of Sentiment Analysis—A Review of Research Topics, Venues, and Top Cited Papers." *Computer Science Review* 27: 16-32.
- [14]. N. Mishra, and C. K. Jha. (2012) "Classification of Opinion Mining Techniques." *International Journal of Computer Applications* 56 (13).
- [15]. Song, Minchae, Hyunjung Park, and Kyung-shik Shin. (2019) "Attention-Based Long Short-Term Memory Network Using Sentiment Lexicon Embedding for Aspect-Level Sentiment Analysis in Korean." *Information Processing & Management* 56 (3): 637-653.
- [16]. P. Sanguansat. (2016, 3-6 Feb. 2016) "Paragraph2Vec-Based Sentiment Analysis on Social Media for Business in Thailand", in the 2016 8th International Conference on Knowledge and Smart Technology (KST).
- [17]. Itani, Maher, Chris Roast, and Samir Al-Khayatt. (2017) "Developing Resources for Sentiment Analysis of Informal Arabic Text in Social Media." *Procedia Computer Science* 117: 129-136.
- [18]. Chekima, Khalifa, and Rayner Alfred. (2018) Sentiment Analysis of Malay Social Media Text. pp. 205-219.
- [19]. D. Cirqueira, M. Fontes Pinheiro, A. Jacob, F. Lobato, and Á. Santana. (2018, 3-6 Dec. 2018). "A Literature Review in Preprocessing for Sentiment Analysis for Brazilian Portuguese Social Media" in the 2018 IEEE/WIC/ACM International Conference on Web Intelligence (WI).
- [20]. Peng, Haiyun, Erik Cambria, and Amir Hussain. (2017) "A Review of Sentiment Analysis Research in the Chinese Language." *Cognitive Computation* 9 (4): 423-435.
- [21]. Ebrahimi, M.m Yazdavar, A., and A. Sheth. (2017) "On the Challenges of Sentiment Analysis for Dynamic Events." *Intelligent Systems, IEEE* 32 (5).
- [22]. Durach, Christian F., Joakim Kembro, and Andreas. (2017) "A New Paradigm for Systematic Literature Reviews in Supply Chain Management." *Journal of Supply Chain Management* Wieland 53 (4): 67-85.
- [23]. Das, Bijoyan, and Sarit Chakraborty. (2018) An Improved Text Sentiment Classification Model Using TF-IDF and Next Word Negation.
- [24]. Khan, Muhammad Taimoor, Mehr Durrani, Armughan Ali, Irum Inayat, Shehzad Khalid, and Kamran Habib Khan. (2016) "Sentiment Analysis and The Complex Natural Language." *Complex Adaptive Systems Modeling* 4 (1): 2.
- [25]. Hassan, Anees Ul, Jamil Hussain, Musarrat Hussain, Muhammad Sadiq, and Sungyoung Lee. (2017) "Sentiment Analysis of Social Networking Sites (SNS) Data Using Machine Learning Approach for the Measurement of Depression", in *International Conference on Information and Communication Technology Convergence (ICTC)*, Jeju, South Korea: IEEE.
- [26]. Akter, Sanjida, and Muhammad Tareq Aziz. (2016) "Sentiment Analysis on Facebook Group Using Lexicon Based Approach", in the 2016 3rd International Conference on Electrical Engineering and Information Communication Technology (ICEEICT).
- [27]. Mahtab, S. Arafin, N. Islam, and M. Mahfuzur Rahaman. (2018, 21-22 Sept. 2018). "Sentiment Analysis on Bangladesh Cricket with Support Vector Machine", in the 2018 International Conference on Bangla Speech and Language Processing (ICBSLP).
- [28]. Dhaoui, Chedia, Cynthia M. Webster, and Lay Peng Tan. (2017) "Social Media Sentiment Analysis: Lexicon Versus Machine Learning." *Journal of Consumer Marketing* 34 (6): 480-488.
- [29]. Rahman, S. A. El, F. A. AlOtaibi, and W. A. AlShehri. (2019, 3-4 April 2019). "Sentiment Analysis of Twitter Data", in the 2019 International Conference on Computer and Information Sciences (ICCIS).
- [30]. Karamollaoglu, H., İ A. Doğru, M. Dörterler, A. Utku, and O. Yıldız. (2018, 20-23 Sept. 2018). "Sentiment Analysis on Turkish Social Media Shares through Lexicon Based Approach", in the 2018 3rd International Conference on Computer Science and Engineering.
- [31]. Mansour, Samah. (2018) "Social Media Analysis of User's Responses to terrorism using sentiment analysis and text mining." *Procedia Computer Science* 140: 95-103.
- [32]. Ali, Kashif, Hai Dong, Athman Bouguettaya, Abdelkarim Erradi, and Rachid Hadjidj. (2017) "Sentiment Analysis as a Service: A Social Media Based Sentiment Analysis Framework", in *IEEE International Conference on Web Services (ICWS)*, Honolulu, HI, USA: IEEE.
- [33]. Ragini, J. Rexiline, P. M. Rubesh Anand, and Vidhyacharan Bhaskar. (2018) "Big Data Analytics for Disaster Response and Recovery Through Sentiment Analysis." *International Journal of Information Management* 42: 13-24.
- [34]. Joyce, Brandon, and Jing Deng. (2017) "Sentiment Analysis of Tweets for the 2016 US Presidential Election", in *IEEE MIT Undergraduate Research Technology Conference (URTC)*, Cambridge, MA, USA: IEEE